Binary Tree Based On-Line Video Summarization

TRECVID 2008 BBC Rushes Summarization Task

Víctor Valdés López
victor.valdes@uam.es

José M. Martínez
josem.martinez@uam.es

Universidad Autónoma de madrid
E28049 Madrid (SPAIN)

Video Processing and Understanding Lab
Grupo de Tratamiento e Interpretación de Vídeo

Index

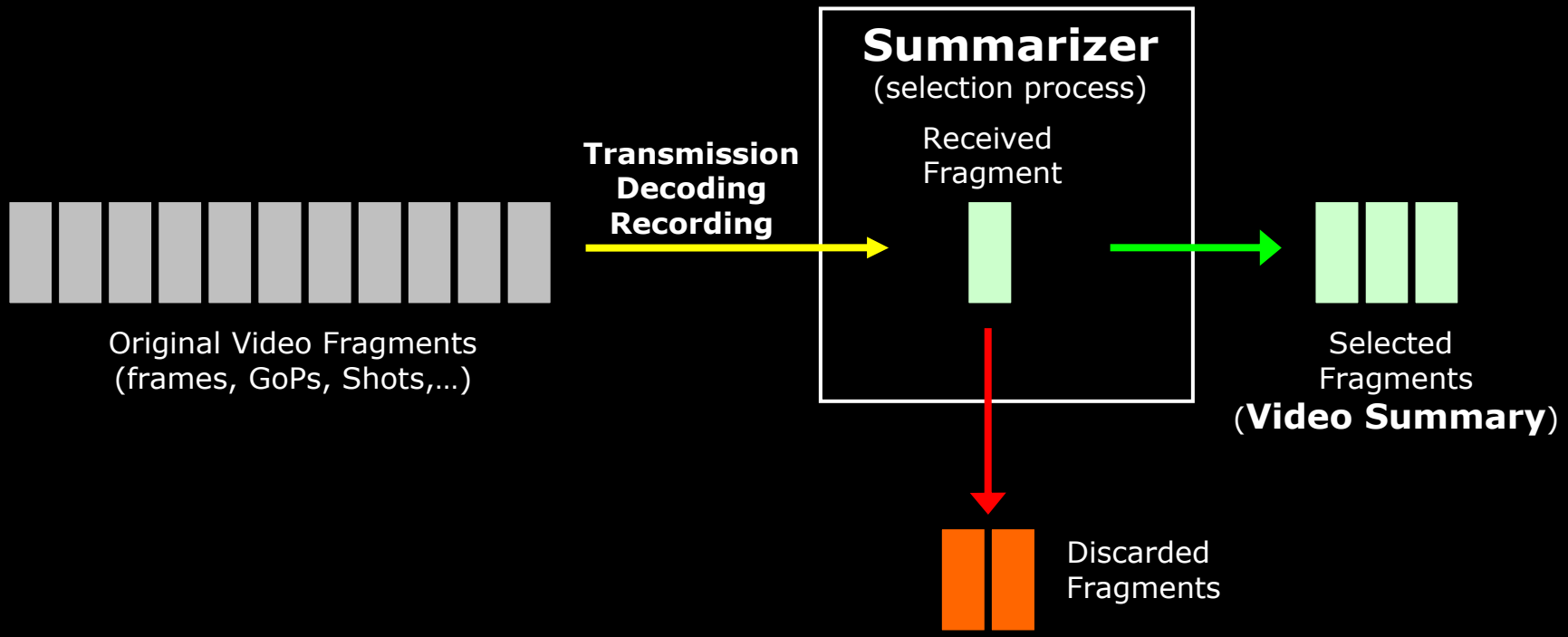
- On-line Video Summarization
  - Introduction
    - Definitions. Off-Line / On-Line Summarization
    - On-line Summarization Pros & Cons.
  - TRECVID 2008 BBC Rushes Summarization Task
    - Summarization Trees
    - Partial Summarization Trees
    - Architecture Overview
    - Summary Scoring
    - Branch Pruning
    - Submitted Runs
    - Results
  - Conclusions

Off-Line / On-Line / Real Time Summarization (I)

- Analysis and Selection Techniques for Video Summarization
  - Off-Line:
    - Video Summarization can have access to the whole available video for the analysis and selection stages.
      - » Clustering
      - » Global Optimization Approaches
      - » …
  - On-Line:
    - Progressive video summary generation (partial video summaries are generated as the original video is received/decoded).
      - » Shot by shot
      - » Frame by frame
      - » GoP by Gop
      - » Others units…
    - Possible real-time summarization if the total summary computation time is shorter than the summary length.

Off-Line / On-Line / Real Time Summarization (II)

- On-Line Video Summarization



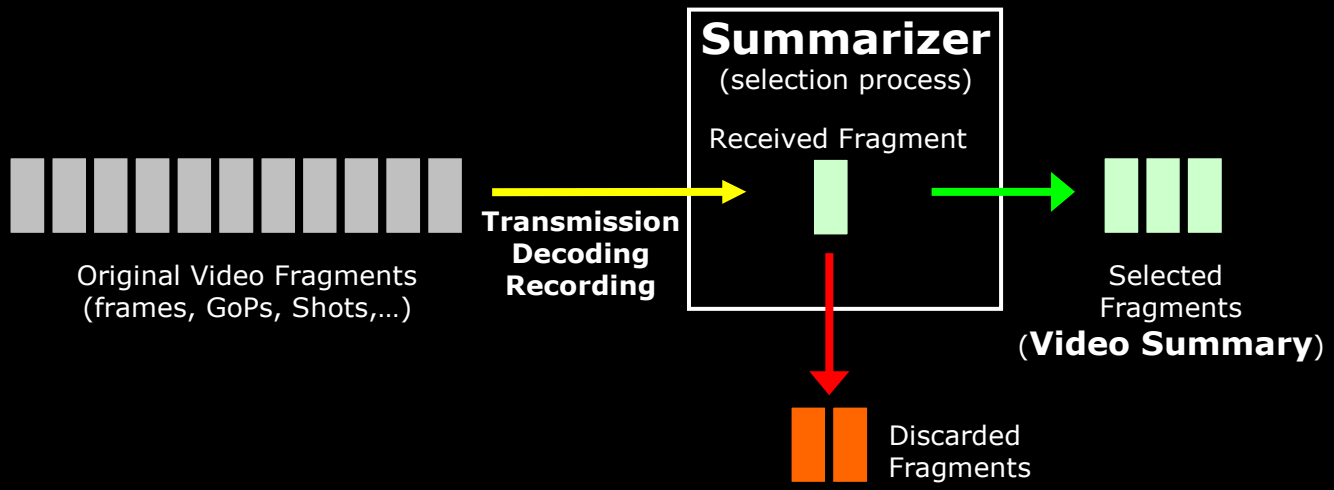
- Summary fragments are selected/discarded before the complete original video is received

Pros & Cons

- On-Line Video Summarization
  - Advantages
    - Provides mechanisms for improving the efficiency of the summarization task.
    - Enables the possibility of:
      - » Generating video summaries of broacasted content as it is being received.
      - » Generating video summaries on recording time.
      - » Watch summaries from stored content on the fly, while being generated.
      - » Creation of interactive summaries.
    - Video summary personalization. For huge video repositories it may be not possible to off-line compute all the summaries necessary to fulfill the specific preferences of each user. Efficient approaches are required.

Pros & Cons

- On-Line Video Summarization
  - Limitations
    - There is not information available about the video that has not been received yet..
    - Once a frame has been written in the output it is not possible to eliminate or substitute it.
    - If the original video length is unknown it can be very difficult to generate a specific-length video summary.

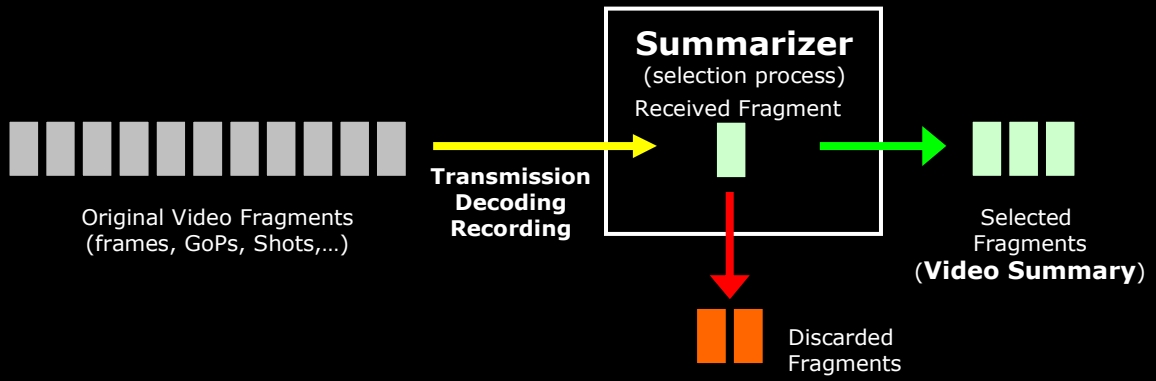


TRECVID 2008 (I)

- TRECVID 2008 – Customizable on-line video summarization framework
  - A novel algorithm for on-line/off-line video summarization has been developed.
  - Customizable in terms of computational complexity, summary generation delay and applied summarization criteria.
  - Previous work focused on ‘instant’ incoming video fragment selection mechanism based on fixed or adaptive thresholds.
  - The selection process can be improved with the possibility of choosing between a number of available video fragments instead of just deciding the inclusion or discardment of each individual video fragment (by introducing a higher delay in the video summary generation).

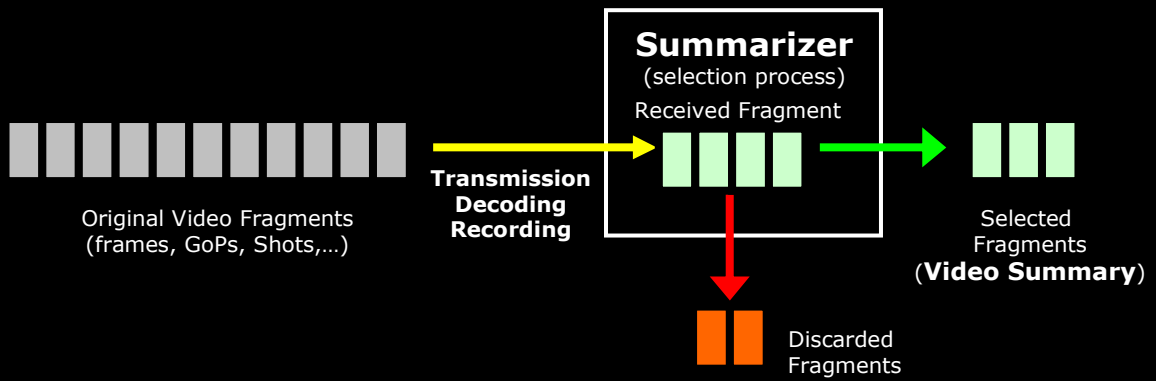
TRECVID 2008 – Summarization Trees (I)

- Instant video fragment selection process:



- Each fragment is selected or discarded as it is received.
- Introduces the smallest possible summarization delay.
- Implies a worse fragment selection.

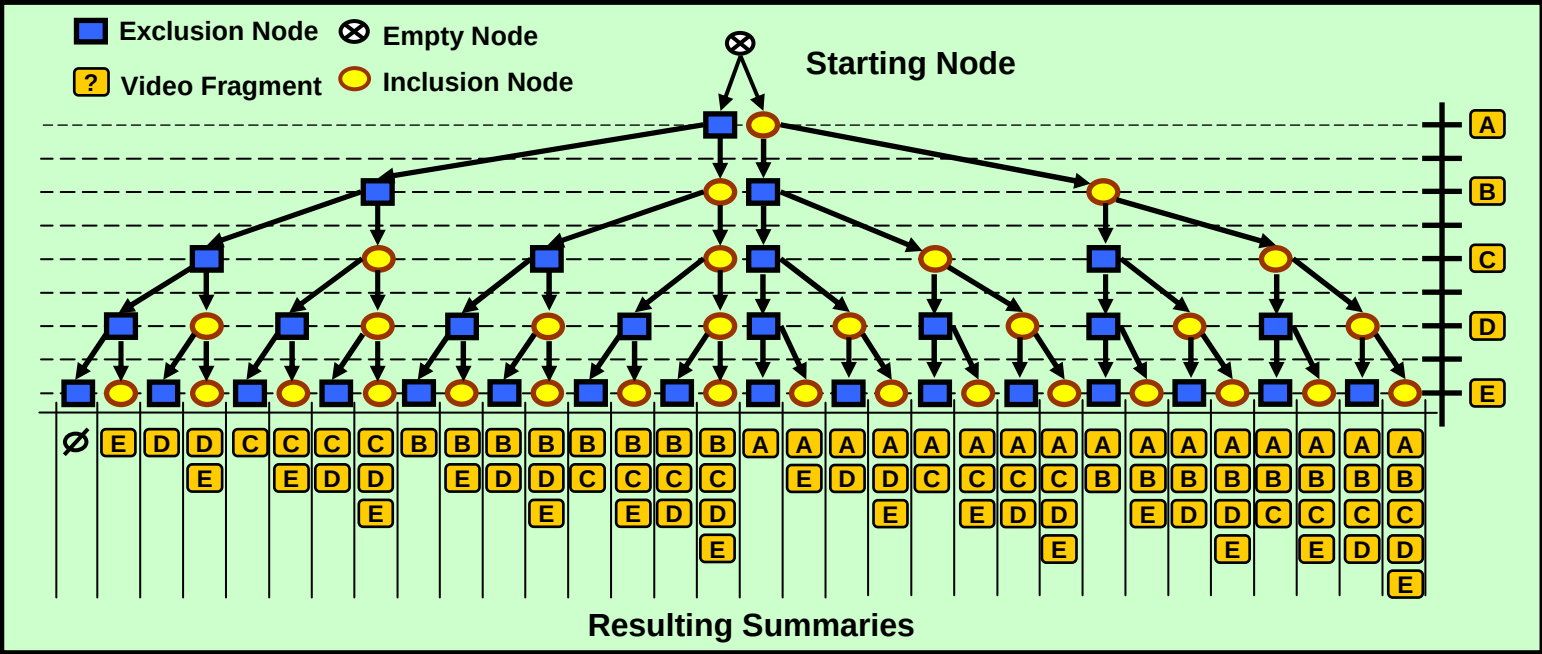
- Buffered video fragment selection process:



- A number of fragments is acumulated before performing selection.
- Introduces higher summarization delay.
- The complexity can increase.
- Enables better fragment selection.

TRECVID 2008 – Summarization Trees (II)

- Summarization trees
  - A ‘summarization tree’ consists on a binary tree which models all the possible video summaries that could be generated from an original video considered as collection of ***n*** basic units (frames, shots, GoPs,etc..).
  - Each node represents the inclusion or discardment of an original video basic unit.
  - Each branch in the tree represents a possible summary (a subset of the original video basic units).

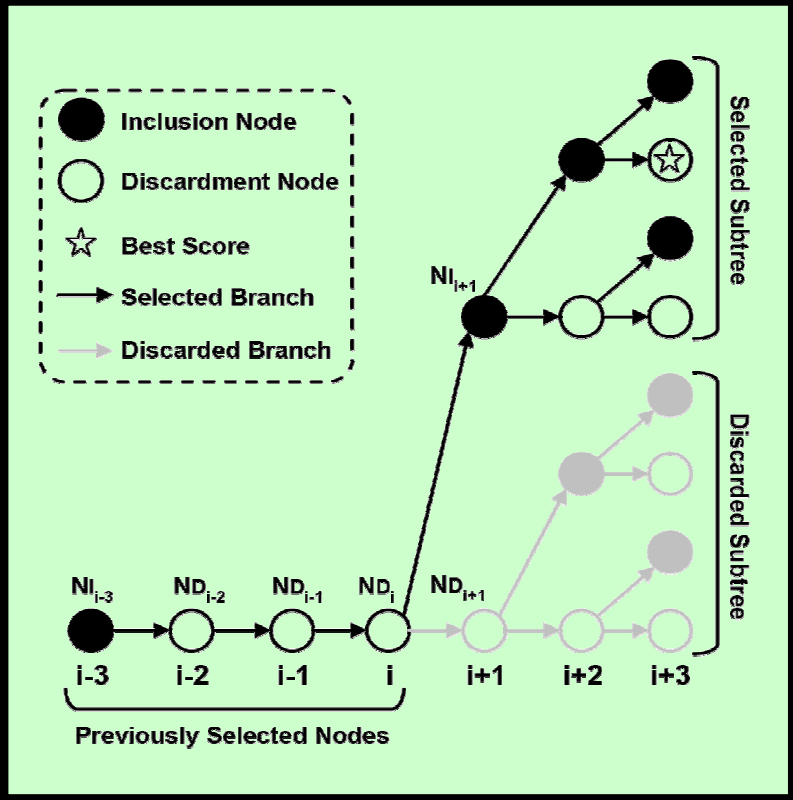


TRECVID 2008 – Summarization Trees (III)

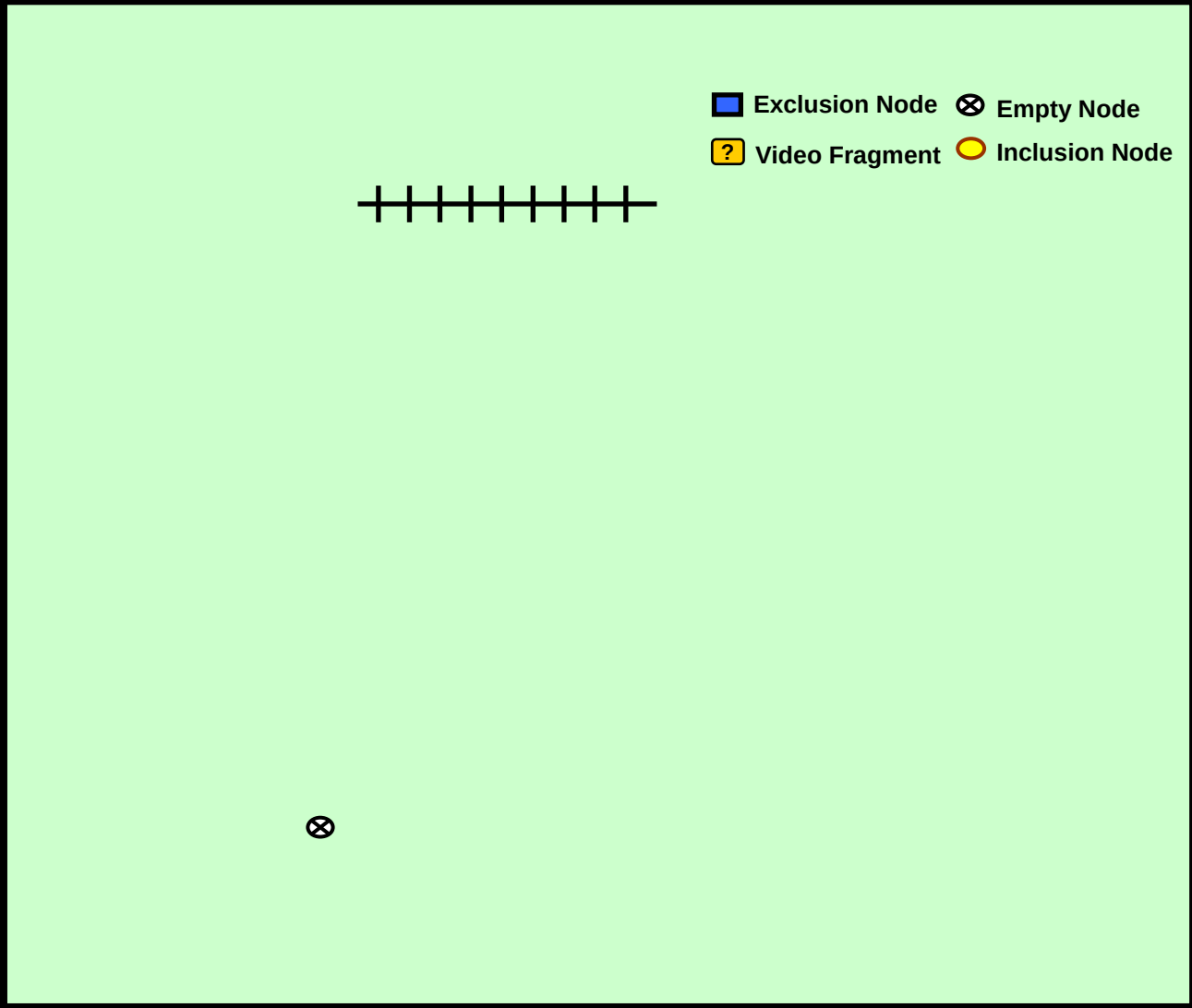
- Summarization trees
  - It is assumed that, for each possible summary (tree branch), a score representing the quality of the video summary can be computed.
  - If all the summaries are represented in the tree and have an associated quality measure it is straightforward to select the best summary given a specific size constraint.
  - **Building a complete tree implies a excesive computational effort**
    - 30 min. Video = 1800 basic units (1 second) = 2^{1800} possible branches.
  - To avoid such high computation effort partial sub-trees will be considered and will allow progressive summary generation.

TRECVID 2008 – Partial Summarization Trees (I)

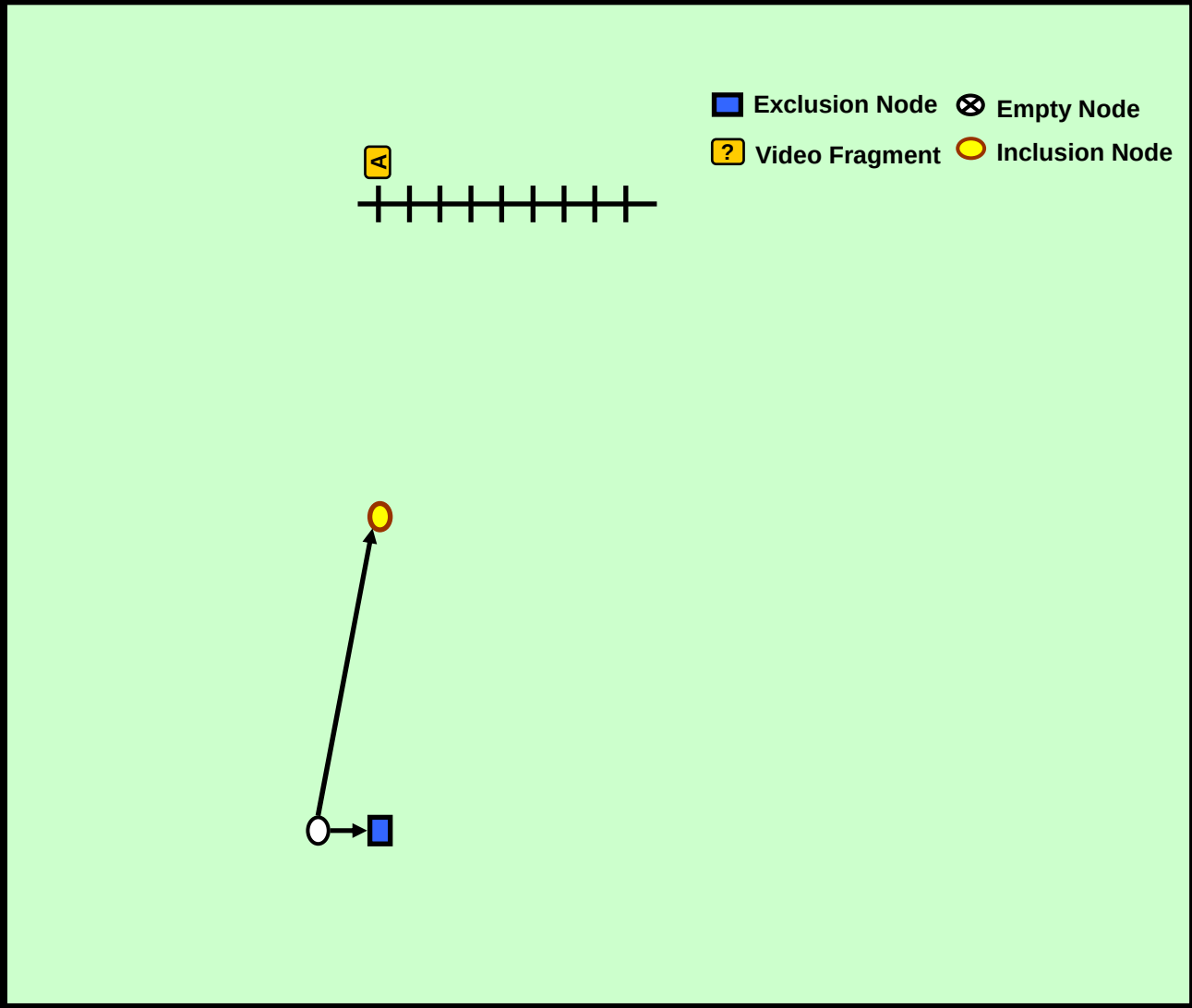
- Partial summarization trees:
  - Only a small summarization tree is kept on memory.
  - Represents the possible combinations of inclusions/discardment of the video fragments not written or discarded yet.
  - Incoming video fragments inclusion/discardment nodes are added to the tree expanding it.
  - Iteratively the subtree including the path with highest score is selected, keeping the size of the summarization tree constant.



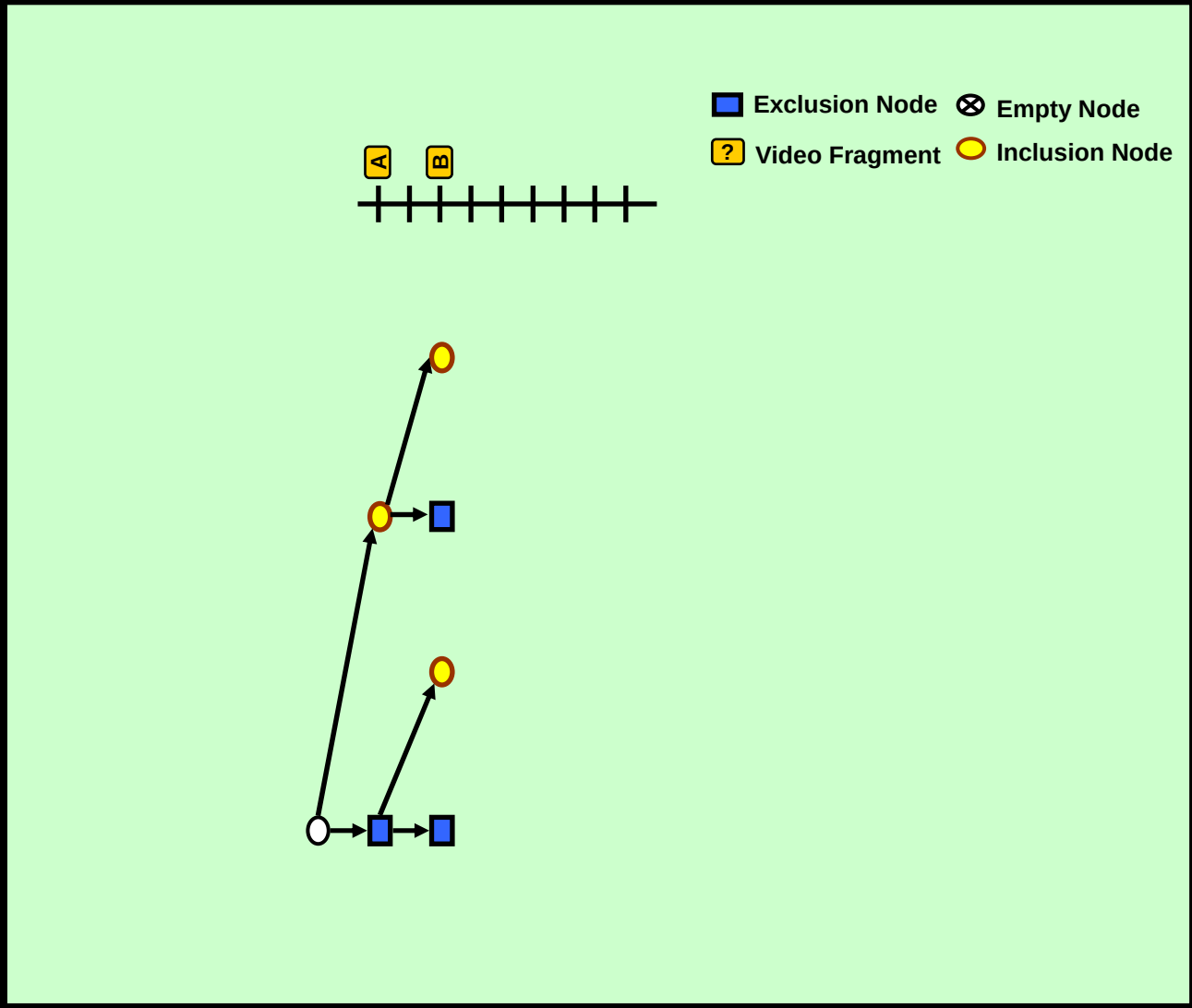
TRECVID 2008 – Partial Summarization Trees (II)



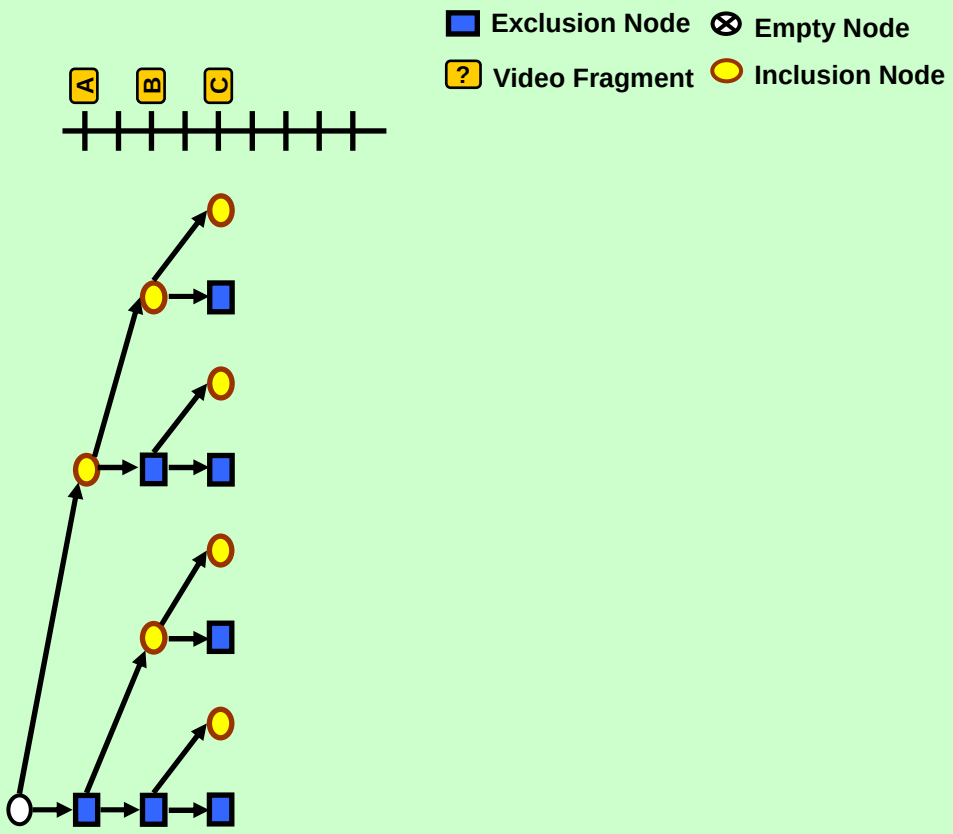
TRECVID 2008 – Partial Summarization Trees (II)



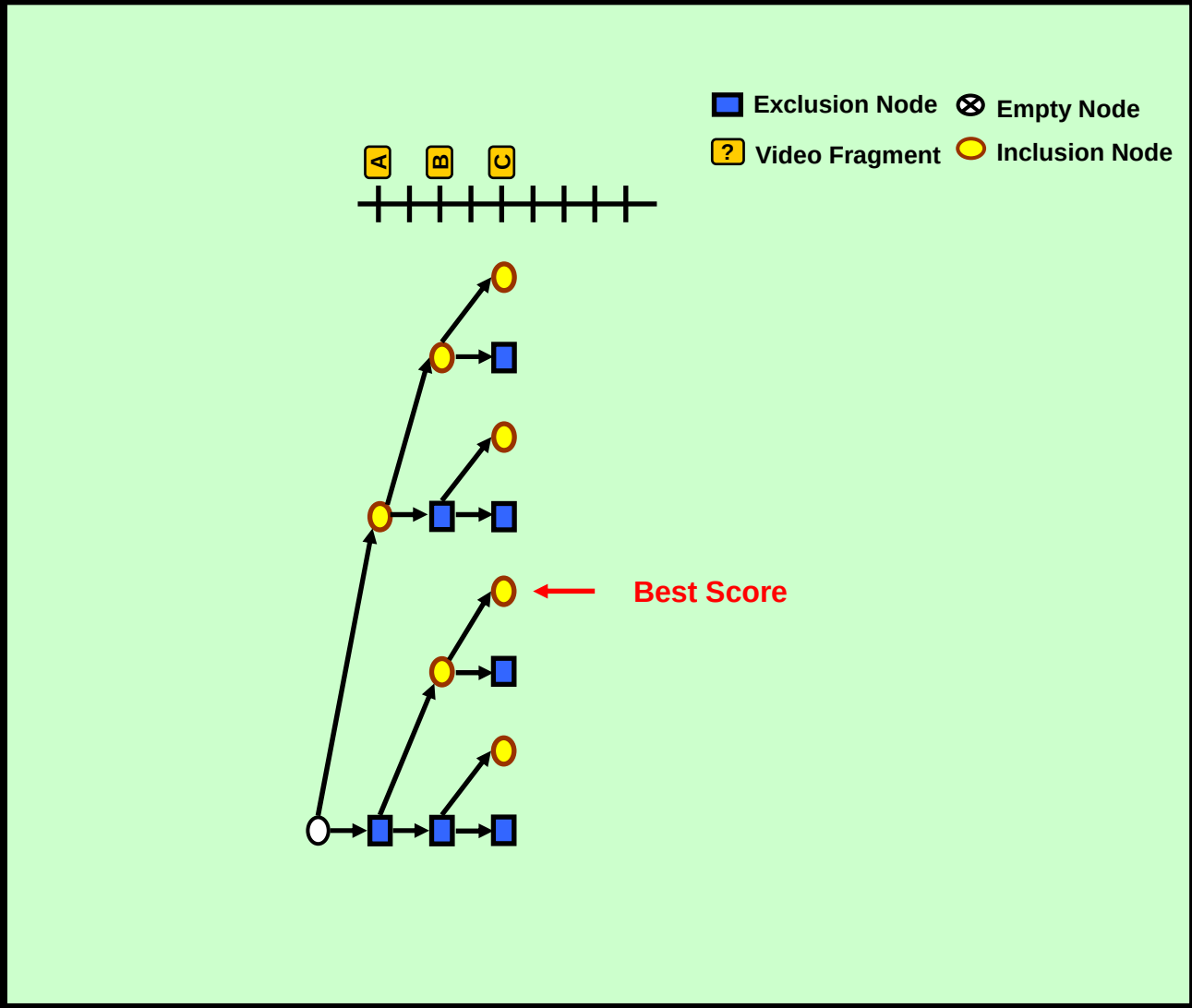
TRECVID 2008 – Partial Summarization Trees (II)



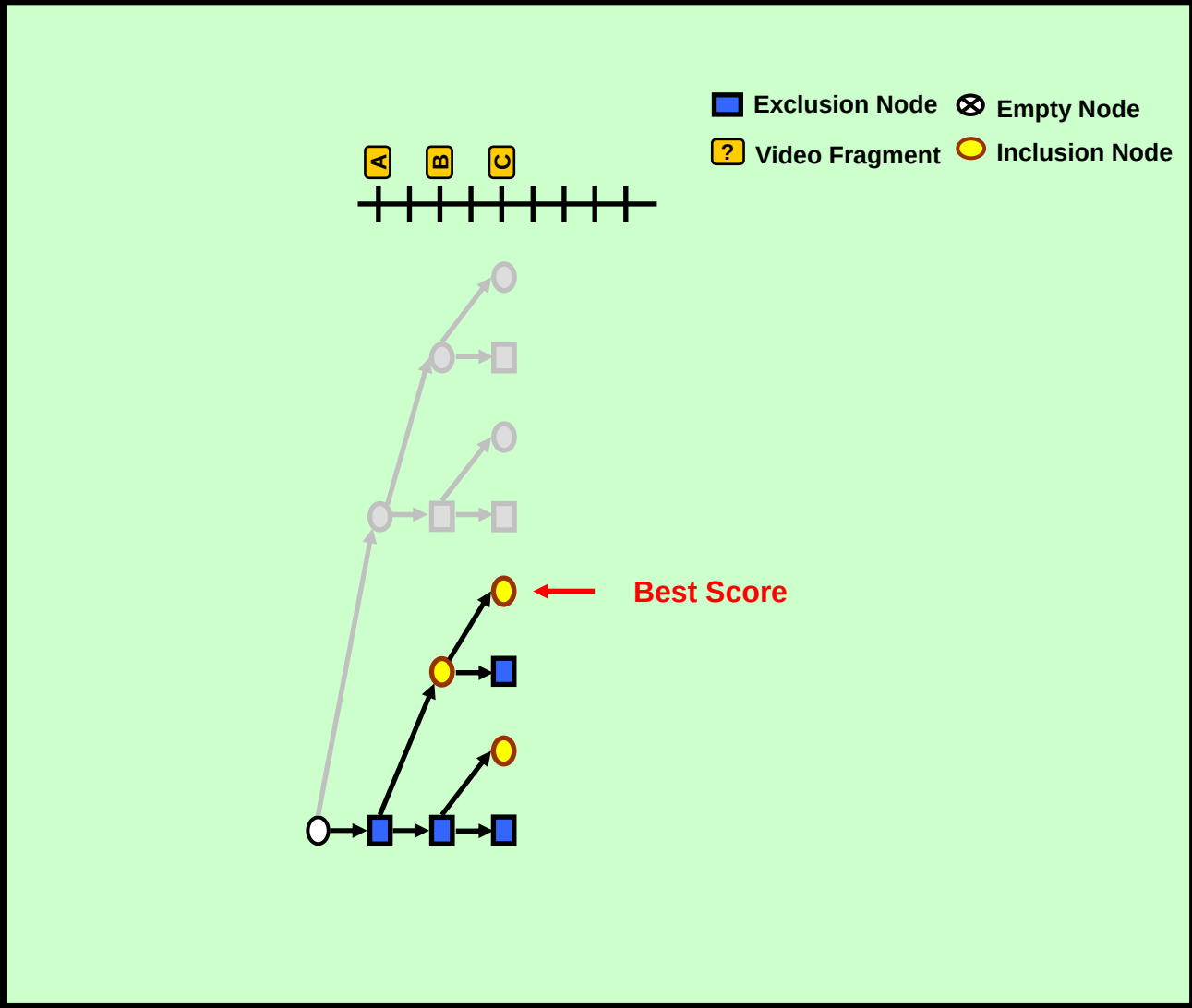
TRECVID 2008 – Partial Summarization Trees (II)



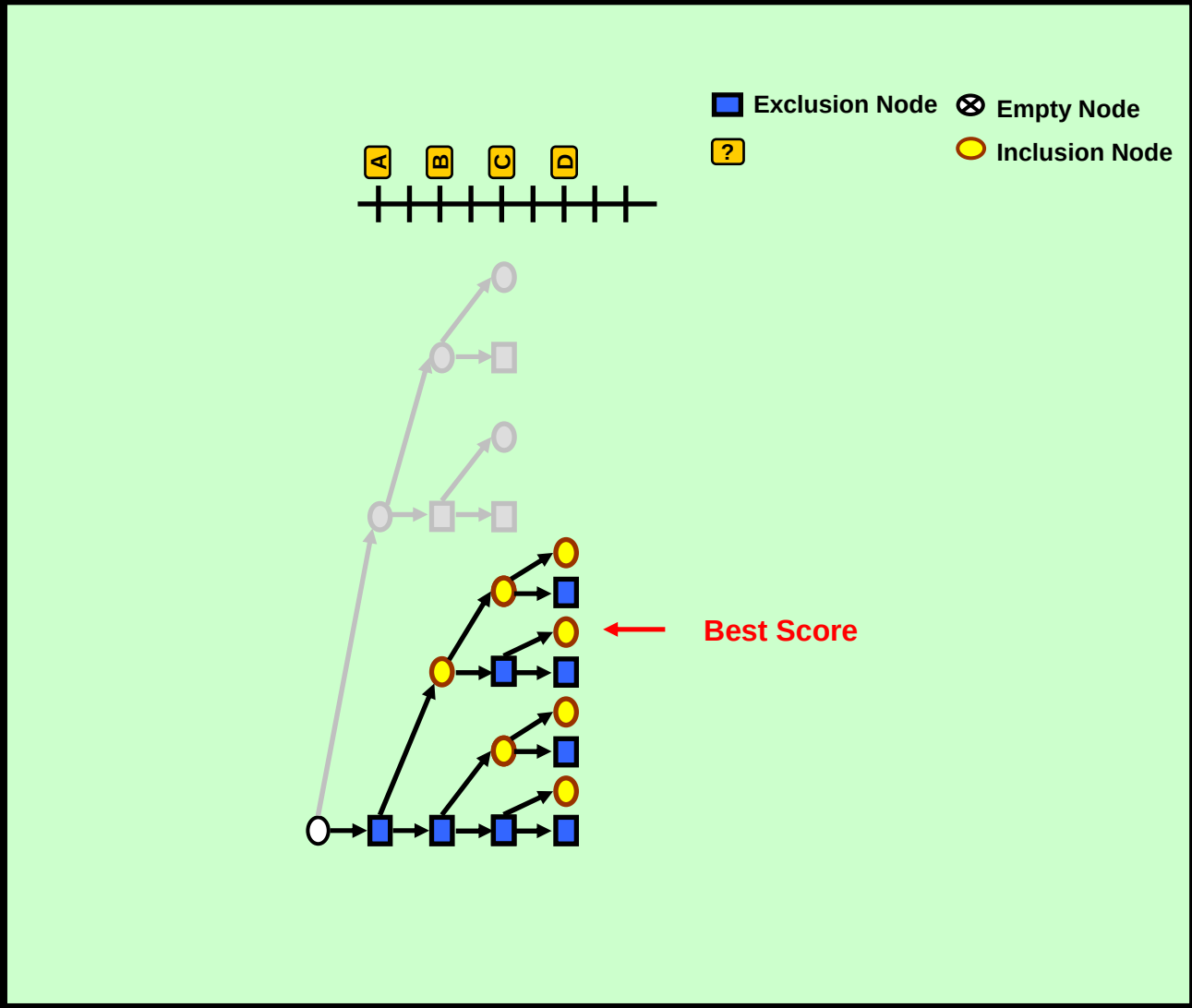
TRECVID 2008 – Partial Summarization Trees (II)



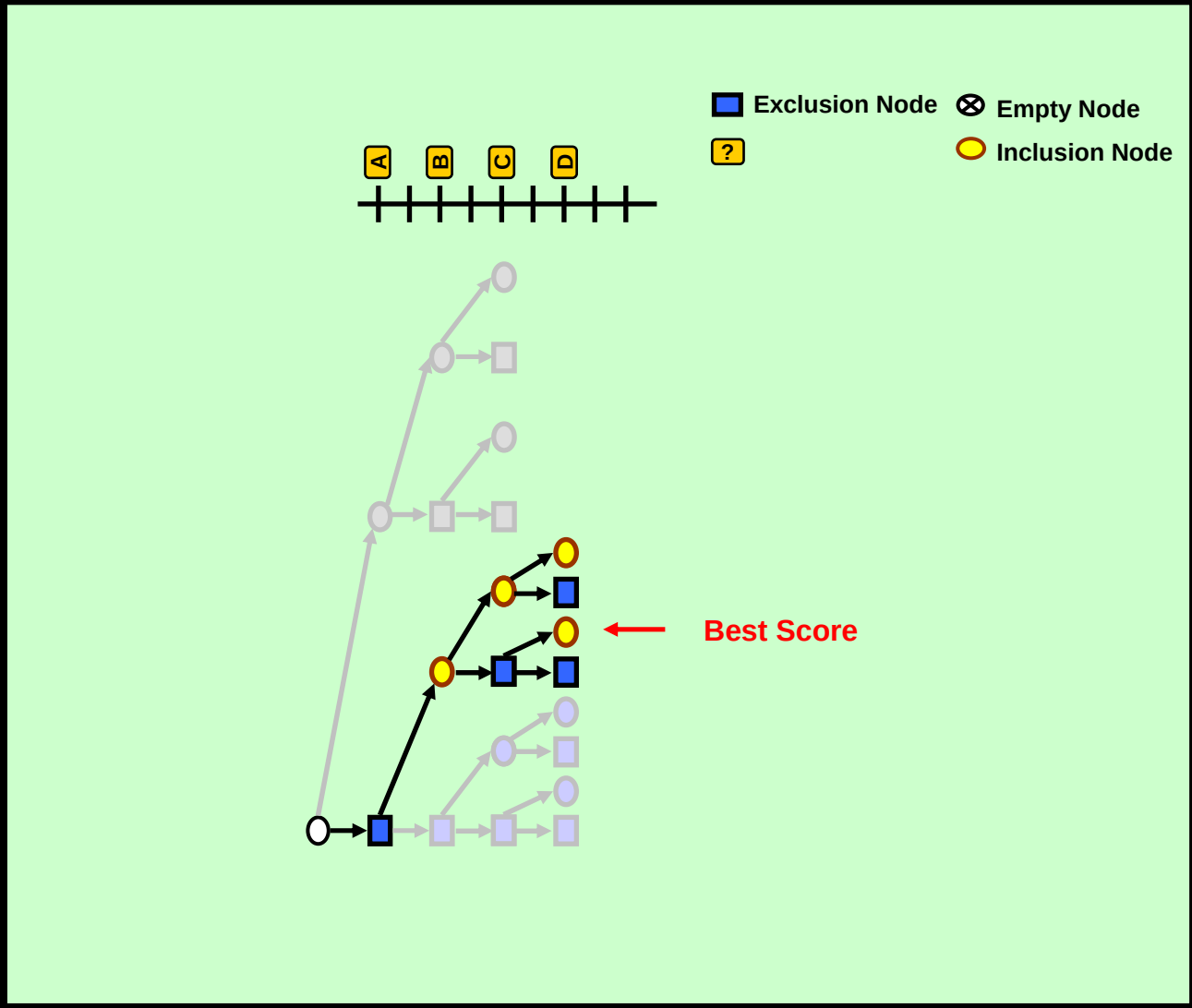
TRECVID 2008 – Partial Summarization Trees (II)



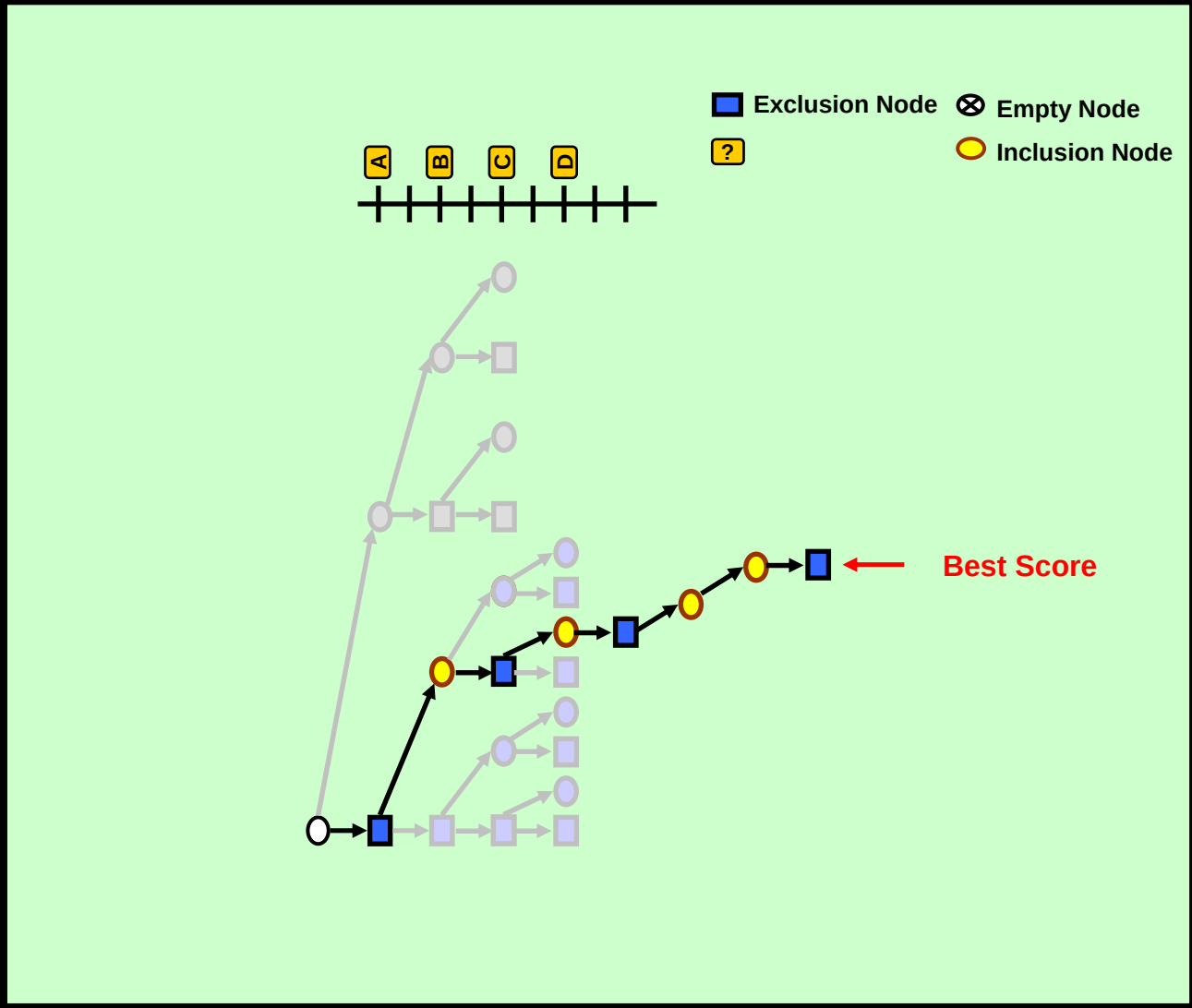
TRECVID 2008 – Partial Summarization Trees (II)



TRECVID 2008 – Partial Summarization Trees (II)



TRECVID 2008 – Partial Summarization Trees (II)



TRECVID 2008 – Partial Summarization Trees (III)

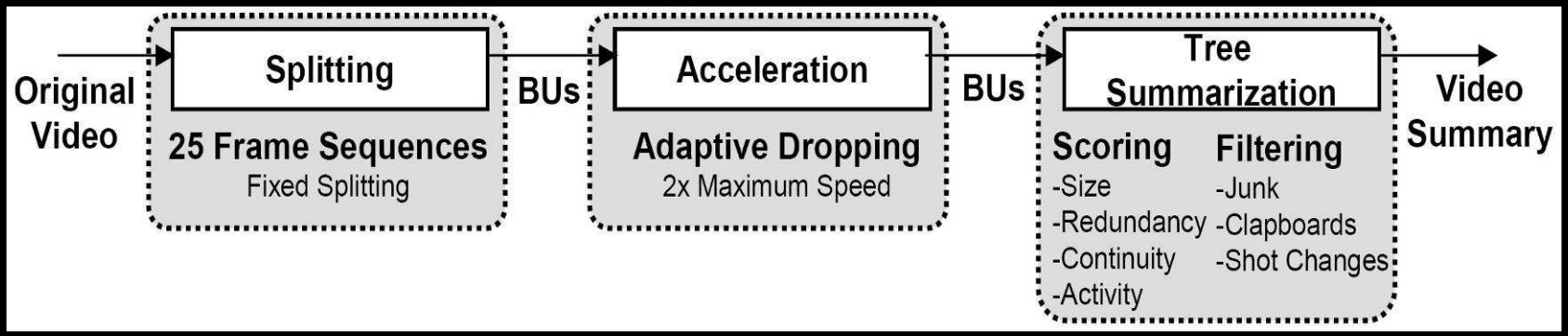
- Summarization Delay Control:
  - Can be controlled by changing the maximum depth of the summarization tree:
    - Higher depth: Better selection results but higher summarization delay and complexity.
    - Lower depth: Reduced summarization delay (even no delay), lower summarization complexity (less possible tree paths on each iteration).
- Computation Effort Control:
  - A high number of branches implies a higher number of path evaluation and computational requirements.
  - Number of branches can be limited to control the summarization effort with a negative impact in the summarization quality.

TRECVID 2008 – Partial Summarization Trees (IV)

- Customizable summary generation:
  - The summary characteristics relies on the branch scoring mechanism, being possible to generate any kind of video summaries (redundancy elimination, higlights, video retrieval…).
- Branch filtering:
  - Easily implementable filtering mechanism just by prunning those branches which include undesired video fragments.

TRECVID 2008 (I): Architecture Overview

- TRECVID 2008 system architecture:
  - Divided in 3 steps:
    - Video splitting: Incoming video is divided in 25 frames (1 second) fixed size fragments which will later feed the tree summarization.
    - Acceleration: A maximum 2x adaptive acceleration is applied. Too similar consecutive frames are dropped (i.e. static shots are shortened).
    - Tree Summarization: A summarization tree, as previously described, is built, obtained a progressive generated video summary.

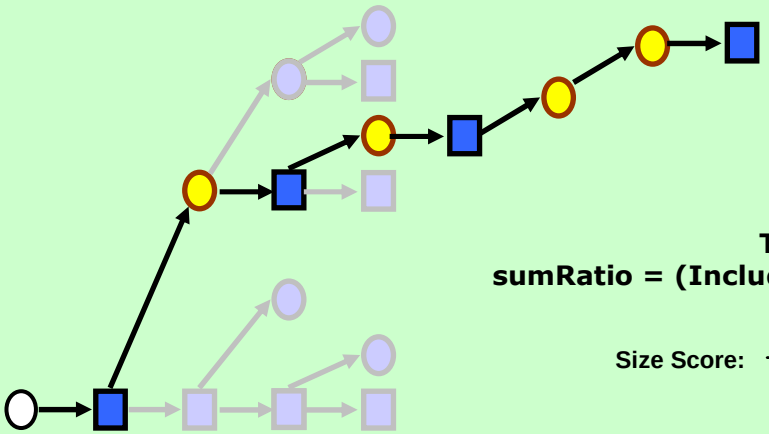


TRECVID 2008 – Summary Scoring (I)

- TRECVID 2008 scoring mechanism:
  - Score based on the combination of several criteria:
    - **Size**
    - **Redundancy**
    - **Activity**
    - **Continuity**
  - Branch pruning applied for avoiding the inclusion of undesired video fragments:
    - **Clapboard detection.**
    - **Junk Filtering.**
    - **Avoiding shot changes.**

TRECVID 2008 – Summary Scoring (II)

- Tree branch scoring mechanism:
  - **Size**: The size of the video summary is limited to 2%. Branches with scores closer to this limit have higher scores.
  - Each node contains information about the number of included and discarded nodes in the branch containing so a summarization ratio and hence a size score can be computed in different ways.



Target Ratio = 0.02

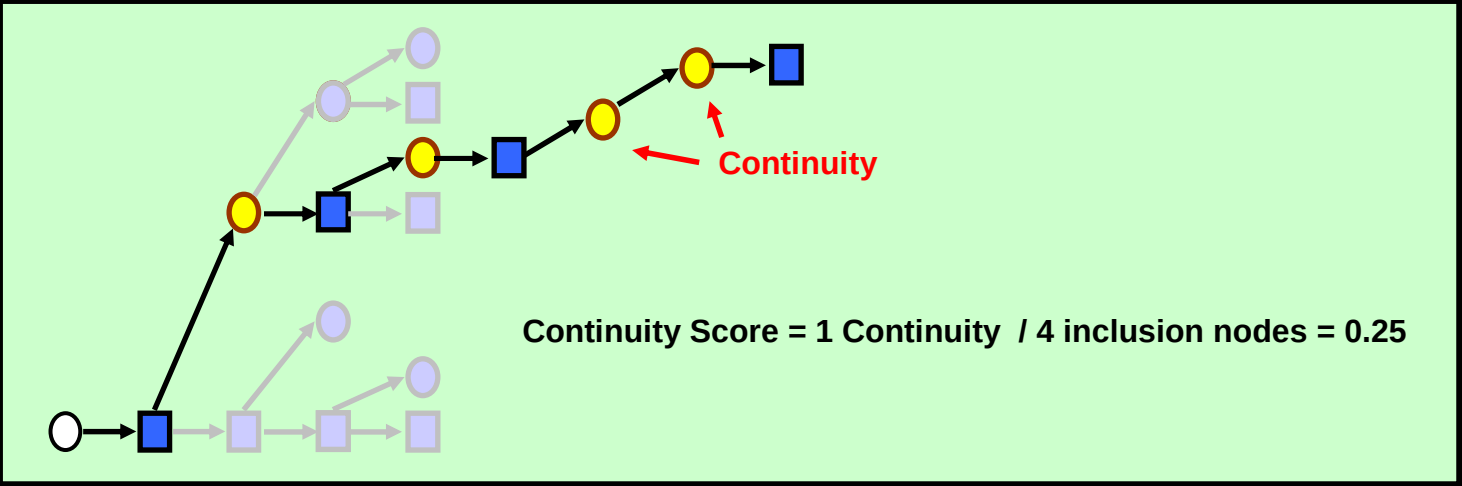
sumRatio = (IncludedNodes/TotalNodes)/Target Ratio

Size Score:

1/sumRatio	if sumRatio > 1
sqrt(sumRatio)	if sumRatio <=1

TRECVID 2008 – Summary Scoring (III)

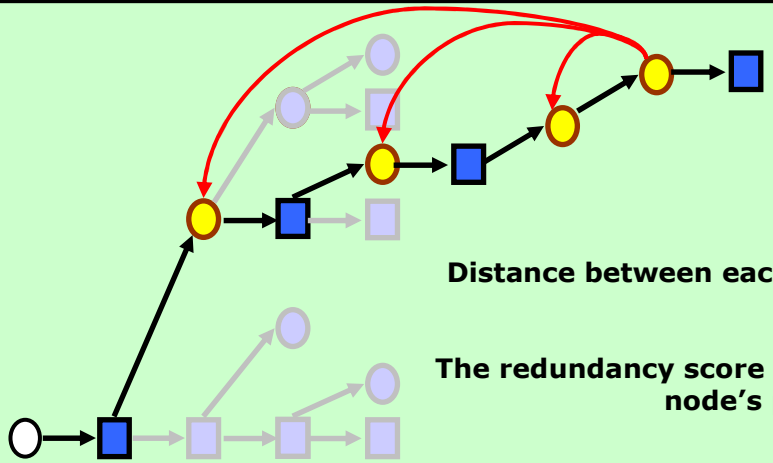
- Tree branch scoring mechanism:
  - **Continuity**: Those tree branches in which the included video fragments are grouped have higher scores, as this result on more pleasant summaries.
  - Continuity score is calculated considering the number of continuities in the video summary (continuities are produced when both a node and its parent are inclusion nodes).



TRECVID 2008 – Summary Scoring (IV)

- Tree branch scoring mechanism:

Shot redundancy: A summary redundancy measure is considered. Each included video fragment is compared with the other fragments included in the branch (making use of the MPEG-7 Color Layout descriptor) obtaining a global summary redundancy measure.

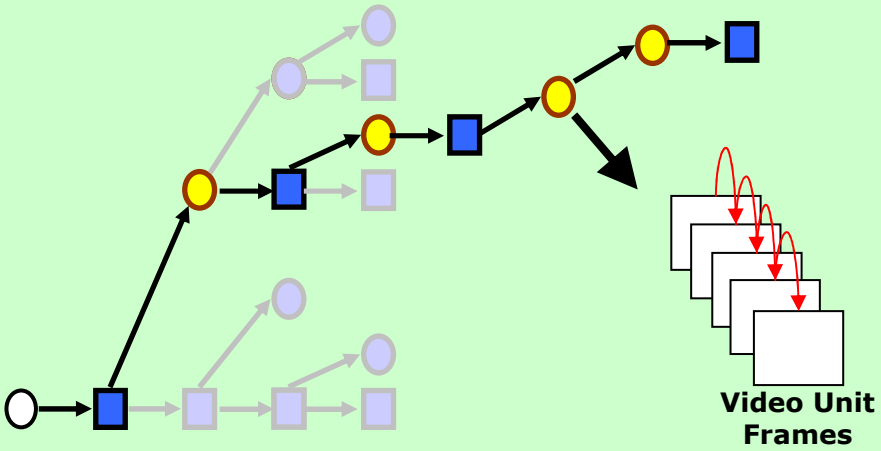


Distance between each node and previously included nodes is averaged.

The redundancy score is iteratively calculated considering each node's parent redundancy score.

TRECVID 2008 – Summary Scoring (V)

- Tree branch scoring mechanism:
  - **Activity:** Non-static video fragments have preference as they are more likely to content video events.



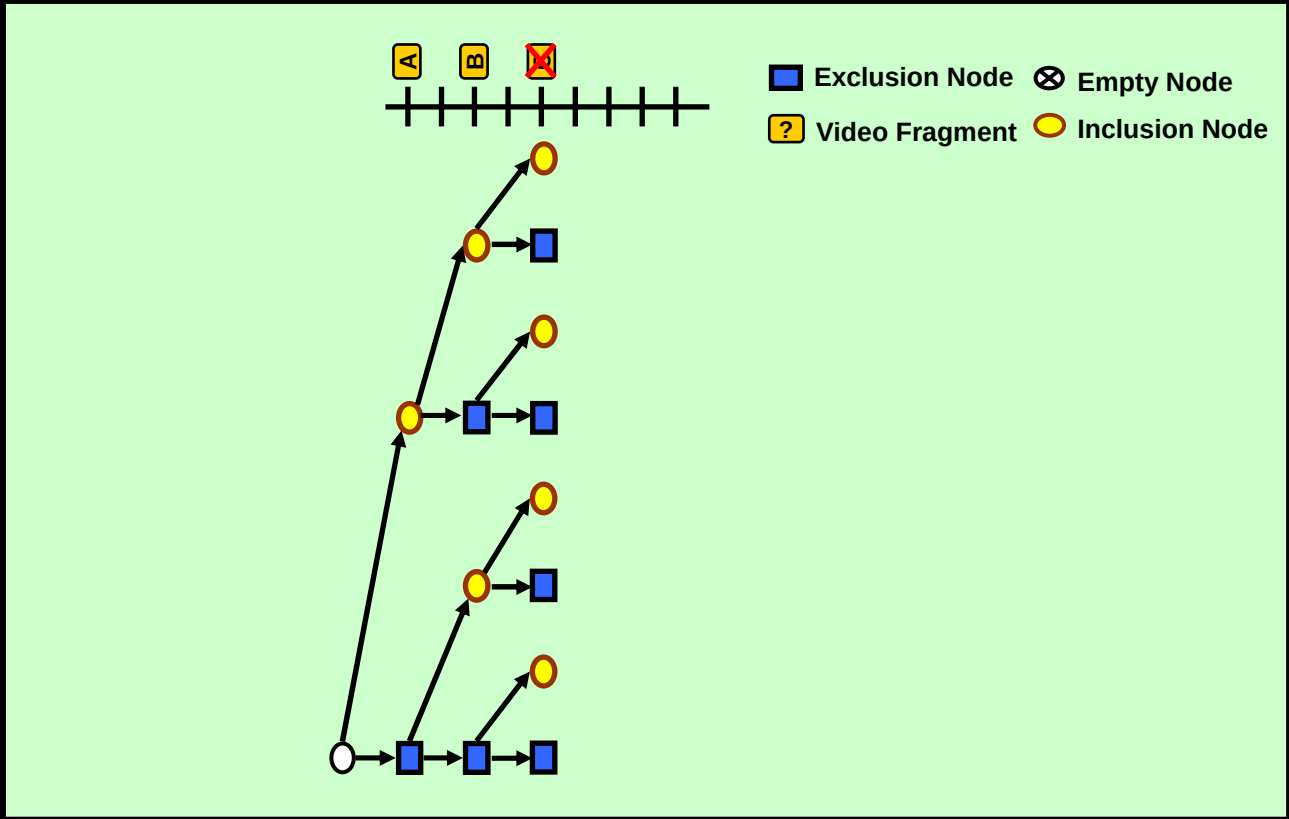
For each video fragment, distance between consecutive frames are averaged, obtaining a video fragment continuity measure.
Activity score is obtained by averaging the activity values for the fragments included in the summary.

TRECVID 2008 – Summary Scoring (VI)

- Tree branch scoring mechanism:
  - Score normalization and combining:
    - The different calculated scores for size, continuity, redundancy and activity are normalized considering the maximum and minimum possible values for each different score in all branches on each tree iteration.
    - For each score a weight is defined: **wSize**, **wContinuity**, **wRedundancy** and **wActivity**. By combining different weights we can obtain summaries in which different characteristics are maximized.
    - The final branch score is calculated by acumulating the weighted score for size, continuity, activity and redundancy.
    - An arbitrary number of scoring functions can be considered and combined in order to get different kind of video summaries.

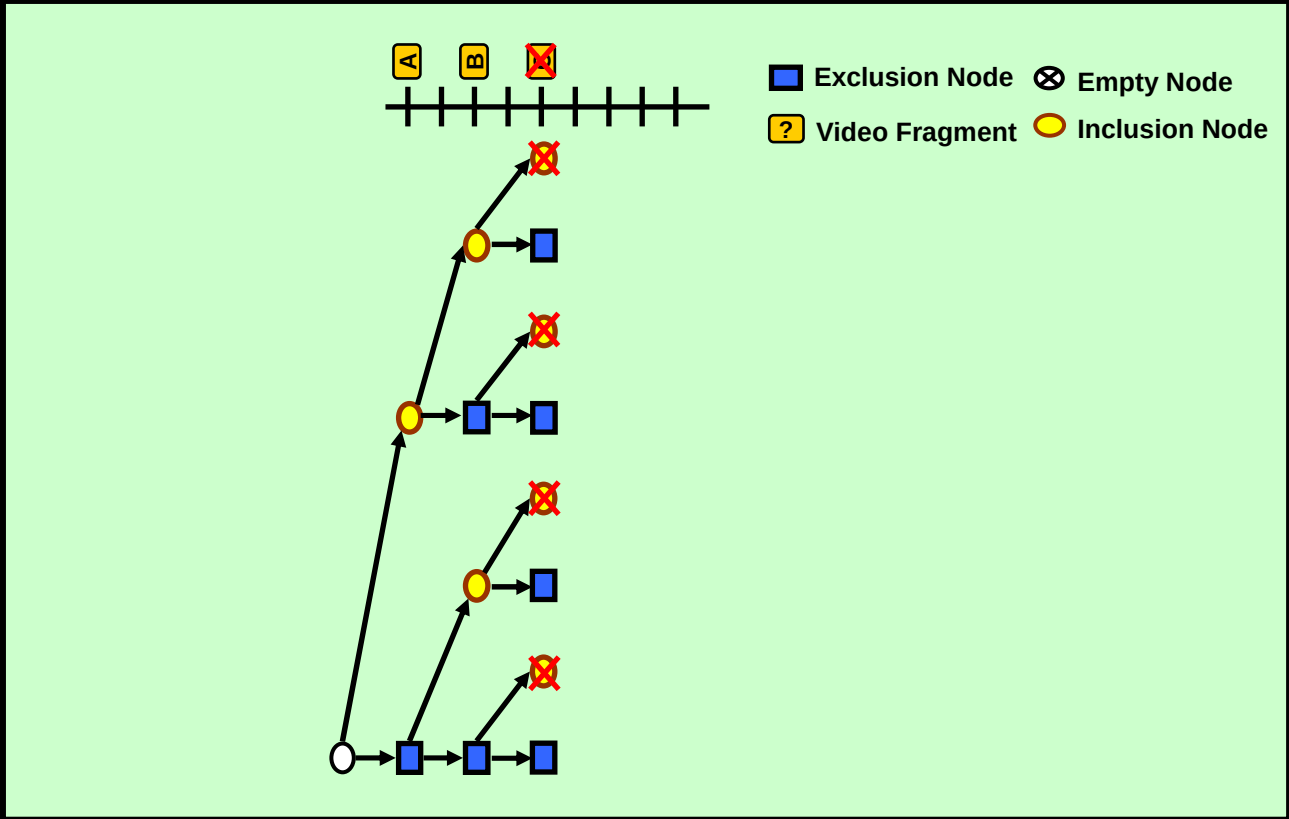
TRECVID 2008 – Branch Filtering (I)

- TRECVID 2008 filtering mechanism:
  - Branch prunning enables to avoid the inclusion of undesired video fragments in the output summaries. Branches including undesired fragments are deleted and only the branches wich do not include those fragments are expanded.



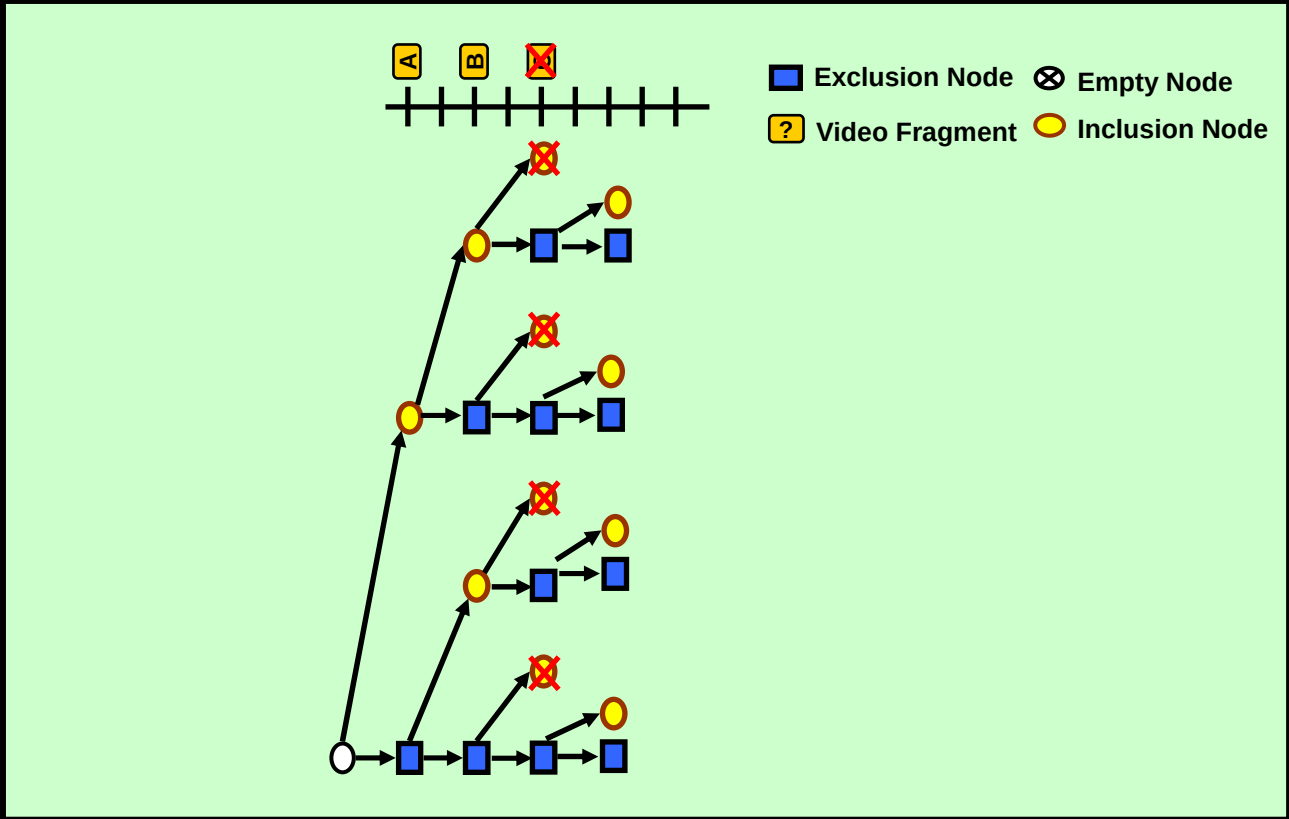
TRECVID 2008 – Branch Filtering (I)

- TRECVID 2008 filtering mechanism:
  - Branch prunning enables to avoid the inclusion of undesired video fragments in the output summaries. Branches including undesired fragments are deleted and only the branches wich do not include those fragments are expanded.



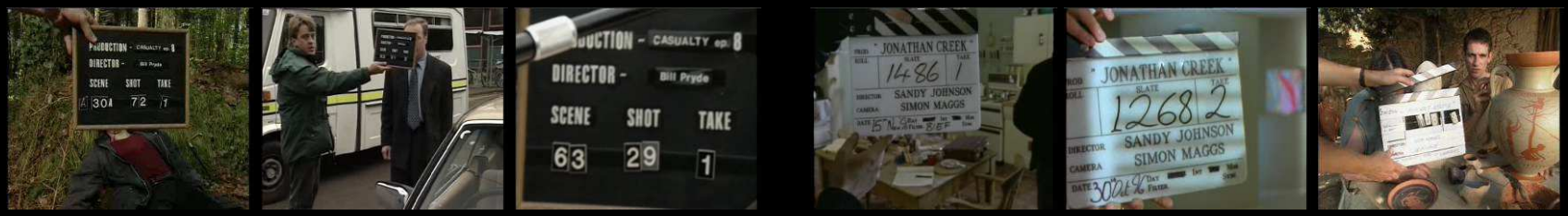
TRECVID 2008 – Branch Filtering (I)

- TRECVID 2008 filtering mechanism:
  - Branch prunning enables to avoid the inclusion of undesired video fragments in the output summaries. Branches including undesired fragments are deleted and only the branches wich do not include those fragments are expanded.



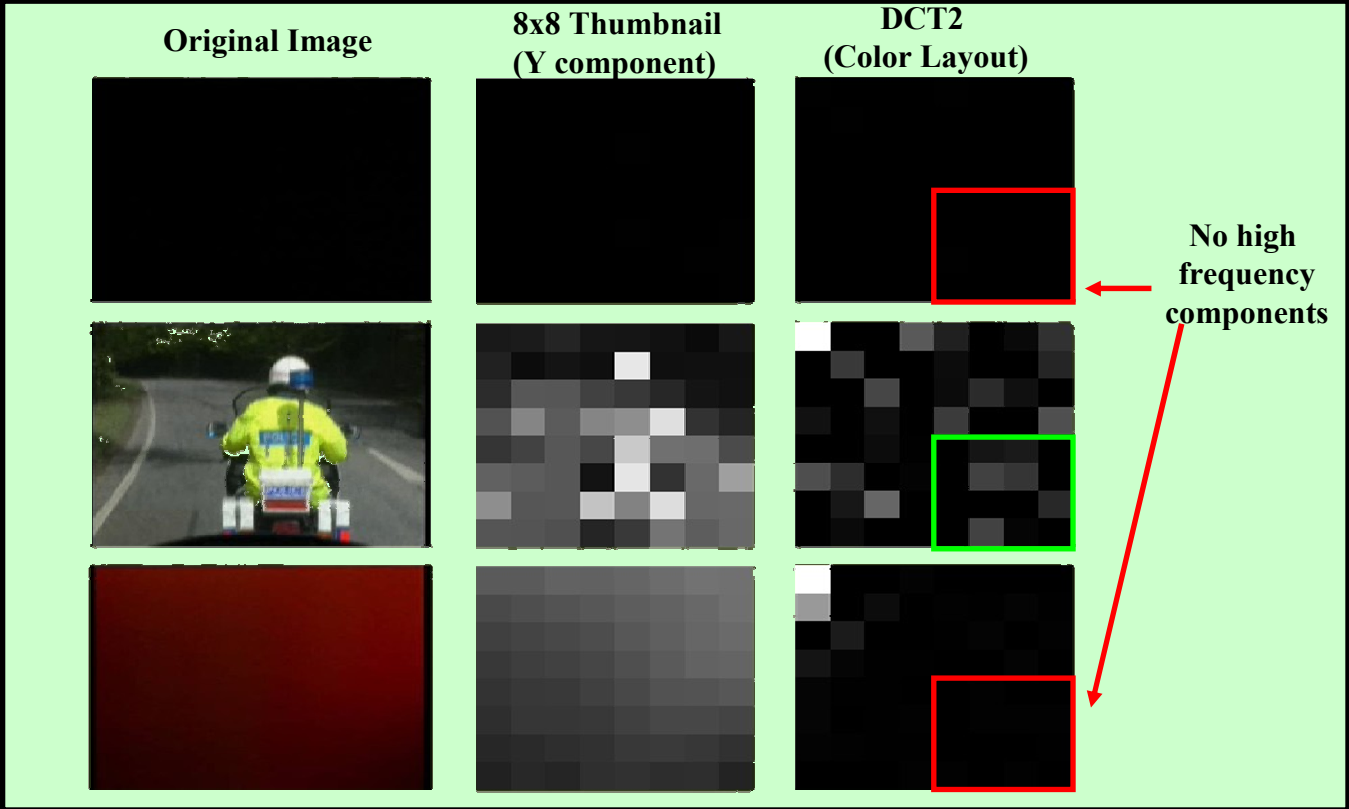
TRECVID 2008 – Branch Filtering (II)

- TRECVID 2008 filtering mechanism:
  - **Clapboard detection:** Fast clapboard detector implemented using openCV haar cascades mechanism. Two different detectors trained for black and white clapboards.
    - » Trained with 1500 examples of white and black clapboards.
    - » Detection rates over 95% obtained over training data.
    - » Results not quantified in the test data set although the high variability in angle, illumination, size and shape of the clapboards reduces the detection rate.



TRECVID 2008 – Branch Filtering (III)

- TRECVID 2008 filtering mechanism:
  - **Junk filter**: Making use of the available 8x8 image thumbnail and DCT (calculated for the MPEG-7 Color Layout) a fast variation measure can be computed to detect blank and test pattern frames.



TRECVID 2008 – Branch Filtering (IV)

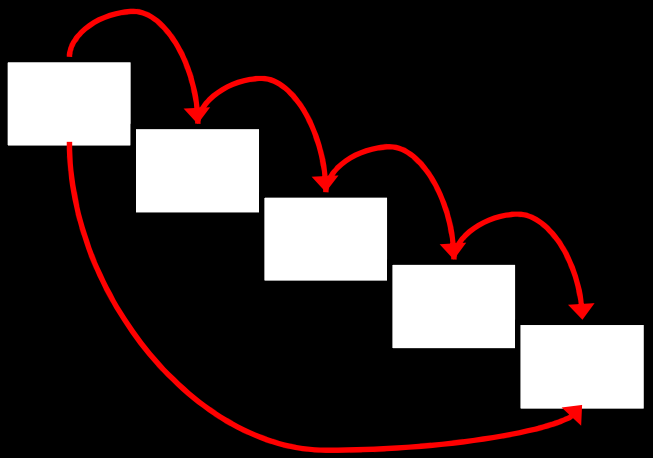
- TRECVID 2008 filtering mechanism:
  - **Junk filter**: Several kind of test patterns can be detected using the image thumbnail just considering that do not contain abrupt vertical color changes.

No vertical variation

Smooth vertical variation

TRECVID 2008 – Branch Filtering (V)

- TRECVID 2008 filtering mechanism:
  - **Shot change**: Video fragments including shot changes can have a negative impact in the video summary smoothness perception.
    - Shot changes are detected by a simple combination between inter-frame comparison between consecutive frames within a video fragment and a comparison between the beginning and end of the video fragment.



Video Fragment Frames

TRECVID 2008 – Submitted Runs (I)

- TRECVID 2008 results:
  - Pentim Xeon 3,7 GHz, 3Gb RAM.
  - Run 1 – Max. Depth=90, max. Branches=1500, wSize=0.475, wRedundancy=0.21, wContinuity=0.12, wVariation=0.195.
    - Avg.Effort: 120 s. (All participants average: 4879 s.)
  - Run 2 – Max. Depth=90, max. Branches=1000, wSize=0.6, wRedundancy=0.35, wContinuity=0.05, wVariation=0.
    - Avg.Effort: 99 s. (All participants average: 4879 s.)

TRECVID 2008 – Submitted Runs (II)

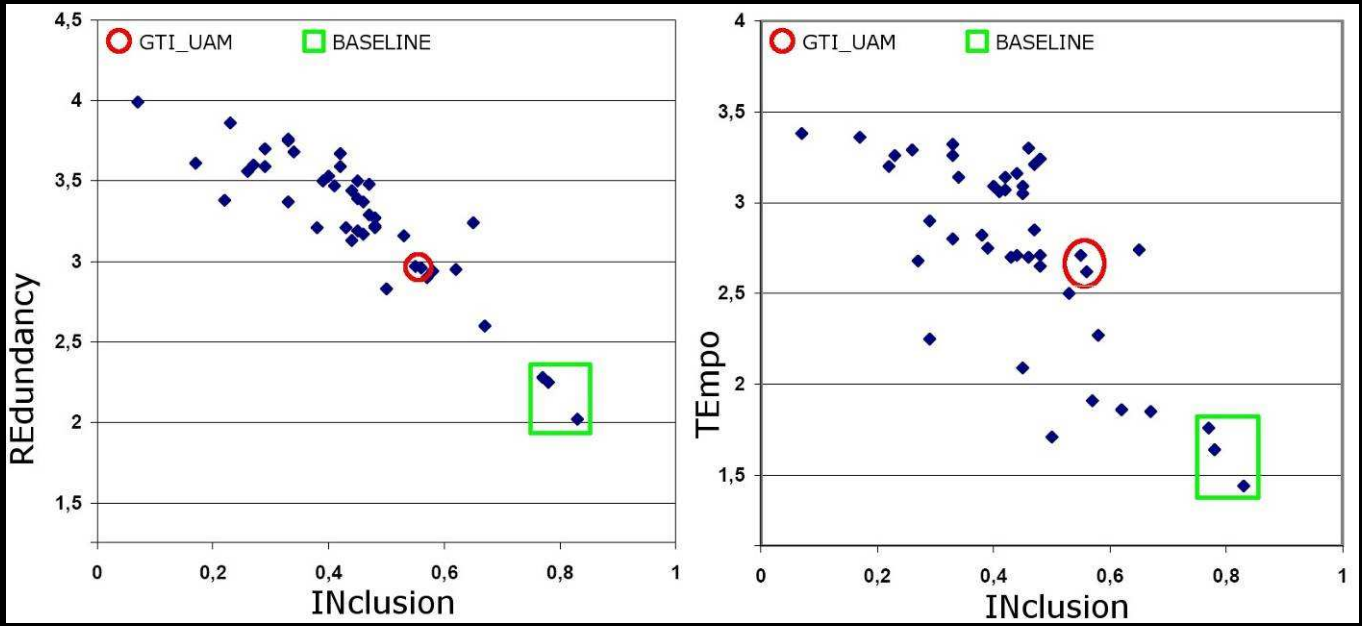
- TRECVID 2008 examples:

Run 1 -MS221050

Run 2 - MS221050

TRECVID 2008 – Results

- TRECVID 2008 results:



Run	DU	XD	TT	VT	IN	JU	RE	TE
Run1	31,2	0,5	45,1	33,1	0,55	3,27	2,97	2,71
Run2	31,1	0,5	47,7	33,5	0,56	3,32	2,96	2,62
Avg.	27	4,5	41,4	29,0	0,46	3,15	3,2	2,72

TRECVID 2008 – Conclusions

- Conclusions:
  - An innovative on-line summarization algorithm has been developed.
  - Provides a generic mechanism for on-line/off-line video summarization in which the performance characteristics of the summarization process can be controlled (speed, summary generation delay).
    - The summarization speed and summary quality are scalable variables:
      - » Faster, low memory consumption but less precise runs or slower high memory consumption runs can be considered.
  - Any scoring mechanism can be implemented using the binary trees method
    - The proposed summary scoring crieteria based on size, redundancy, continuity and activity with junk filtering mechanisms have proved to work well with BBC rushes content.
  - On-Line approache enables new applications in the future (summaries played on generation time, summary personalization, interactivity).

References

- Valdes, V., Martínez J.M. “Binary Tree Based On-Line Video Summarization”, Proceedings of the ACM Multimedia 2008 (Workshop on TRECVID video Summarization), Vancouver, Canada, 27 Octobe –1 November 2008, pp.134-138.

- Previous Work:

  - Valdes, V., Martínez J.M. “On-Line Video Summarization Based on Signature-Based Junk and Redundancy Filtering”, Proceedings of the Ninth International Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS’08, pp 88-91, 7-9 May 2008.

  - Valdes, V., Martínez J.M. “Post-Processing Techniques for On-line Adaptive Video Summarization Based on Relevance Curves”, Semantic Media and Digital Media Technologies – SAMT’07, Lecture Notes in Computer Science, Vol. 4816, pp: 144-157, Springer Verlag, Dec. 2007

  - Valdes, V., Martínez J.M. “On-Line Video Skimming Based on Histogram Similarity”, Proceedings of the ACM Multimedia 2007, Workshop on TRECVID Video Summarization, Augsburg, Germany, 24-29 September 2007, pp. 94-98.

Thank you for your attention!